

Derivačné stromy, redukovaná gramatika

Ing. Viliam Hromada, PhD.

C-510
Ústav informatiky a matematiky
FEI STU

`viliam.hromada@stuba.sk`



Bezkontextová gramatika č. 1

Je daná gramatika $G = (N, T, P, S)$. $N = \{S, A, B\}$, $T = \{a, b\}$. Pravidlá:

- $S \rightarrow abS \mid AB$
- $A \rightarrow a \mid aA \mid aBa$
- $B \rightarrow b \mid bS$

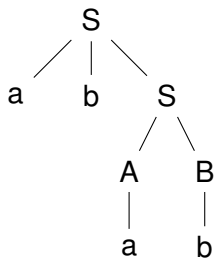
Pre uvedenú gramatiku splňte nasledovné úlohy:

1. Nájdite derivácie reťazcov $abab$, $aabab$ a nakreslite ich derivačné stromy.
2. Zostrojte ľavé a pravé derivácie reťazcov $abab$, $aabab$.
3. Je daná gramatika jednoznačná?



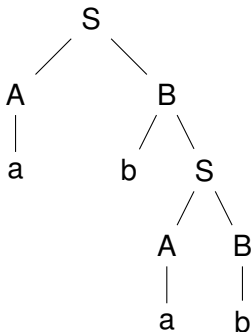
Derivácia *abab* č. 1:

$S \Rightarrow abS \Rightarrow abAB \Rightarrow abaB \Rightarrow abab$



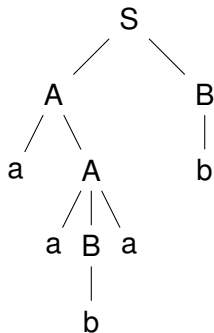
Derivácia *abab* č. 2:

$S \Rightarrow AB \Rightarrow aB \Rightarrow abS \Rightarrow abAB \Rightarrow abaB \Rightarrow abab$



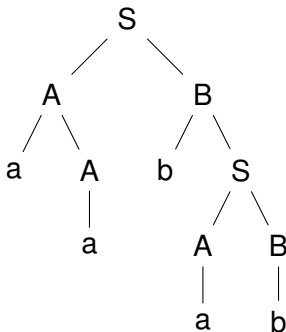
Derivácia *aabab* č. 1:

$S \Rightarrow AB \Rightarrow aAB \Rightarrow aaBaB \Rightarrow aabaB \Rightarrow aabab$



Derivácia *aabab* č. 2:

$S \Rightarrow AB \Rightarrow aAB \Rightarrow aaB \Rightarrow aabS \Rightarrow aabAB \Rightarrow aabaB \Rightarrow aabab$



Pre obe derivácie $abab$ vieme uvažovať ich ľavé (pravé) verzie: Ľavá derivácia $abab$ z derivácie č. 1:

$$S \Rightarrow_l abS \Rightarrow_l abAB \Rightarrow_l abaB \Rightarrow_l abab$$

Pravá derivácia $abab$ z derivácie č. 1:

$$S \Rightarrow_r abS \Rightarrow_r abAB \Rightarrow_r abAb \Rightarrow_r abab$$



Ľavá derivácia *abab* z derivácie č. 2:

$$S \Rightarrow_l AB \Rightarrow_l aB \Rightarrow_l abS \Rightarrow_l abAB \Rightarrow_l abaB \Rightarrow_l abab$$

Pravá derivácia *abab* z derivácie č. 2:

$$S \Rightarrow_r AB \Rightarrow_r AbS \Rightarrow_r AbAB \Rightarrow_r AbAb \Rightarrow_r Abab \Rightarrow_r abab$$

Podobne pre obe derivácie *aabab* vieme uvažovať ich ľavé (pravé) verzie: Ľavá derivácia *aabab* z derivácie č. 1:

$$S \Rightarrow_l AB \Rightarrow_l aAB \Rightarrow_l aaBaB \Rightarrow_l aabaB \Rightarrow_l aabab$$

Pravá derivácia *aabab* z derivácie č. 1:

$$S \Rightarrow_r AB \Rightarrow_r Ab \Rightarrow_r aAb \Rightarrow_r aaBab \Rightarrow_r aabab$$



Ľavá derivácia *aabab* z derivácie č. 2:

$$S \Rightarrow_l AB \Rightarrow_l aAB \Rightarrow_l aaB \Rightarrow_l aabS \Rightarrow_l aabAB \Rightarrow_l aabaB \Rightarrow_l aabab$$

Pravá derivácia *aabab* z derivácie č. 2:

$$S \Rightarrow_r AB \Rightarrow_r AbS \Rightarrow_r AbAB \Rightarrow_r AbAb \Rightarrow_r Abab \Rightarrow_r aAbab \Rightarrow_r aabab$$



Zároveň sa nám podarilo zodpovedať otázku, či je gramatika jednoznačná:

- Gramatika **nie je jednoznačná**, teda je **nejednoznačná**, pretože
- **Existuje reťazec**, ktorý má aspoň 2 rôzne derivačné stromy - dokonca sme zistili, že existujú minimálne 2 reťazce: *abab* alebo *aabab*, pretože v oboch prípadoch platí, že majú minimálne 2 rôzne derivačné stromy.
- Napr. pre reťazec *abab* vidno, že derivačný strom na slajde č. 3 je **iný** než derivačný strom na slajde č. 4.



Bezkontextová gramatika č. 2

Je daná gramatika $G = (N, T, P, S)$. $N = \{S, A, B\}$, $T = \{a, b\}$. Pravidlá:

- $S \rightarrow AaB \mid BbA$
- $A \rightarrow bAa \mid \varepsilon$
- $B \rightarrow b \mid S$

Dokážte pre uvedenú gramatiku že nie je jednoznačná tým, že nájdete 2 rôzne derivačné stromy pre reťazec *baabb*:



Táto gramatika je **nejednoznačná**, pretože pre reťazec *baabb* existuje derivácia:

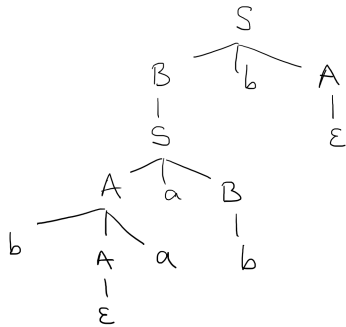
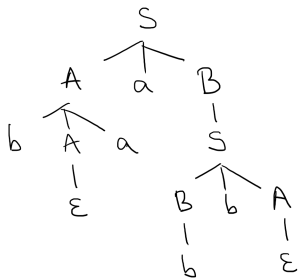
$$S \Rightarrow AaB \Rightarrow bAaaB \Rightarrow baaB \Rightarrow baaS \Rightarrow baaBbA \Rightarrow baabbA \Rightarrow baabb$$

a rovnako existuje iná derivácia:

$$S \Rightarrow BbA \Rightarrow SbA \Rightarrow AaBbA \Rightarrow bAaaBbA \Rightarrow baaBbA \Rightarrow baabbA \Rightarrow baabb$$

pričom **derivačné stromy** oboch derivácií sú rôzne - vid' ďalší slajd.





Bezkontextová gramatika č. 3

Je daná bezkontextová gramatika s pravidly:

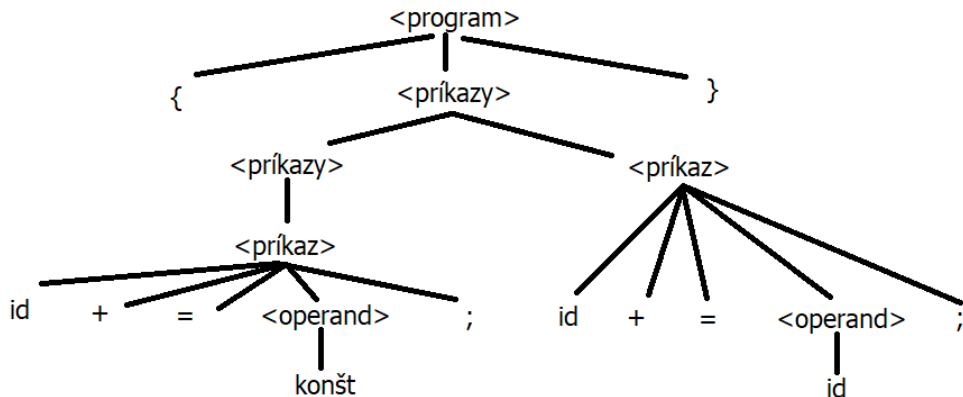
1. $\langle \text{program} \rangle \rightarrow \{ \langle \text{příkazy} \rangle \}$
2. $\langle \text{příkazy} \rangle \rightarrow \langle \text{příkaz} \rangle$
3. $\langle \text{příkazy} \rangle \rightarrow \langle \text{příkazy} \rangle \langle \text{příkaz} \rangle$
4. $\langle \text{příkaz} \rangle \rightarrow \mathbf{id += \langle \text{operand} \rangle ;}$
5. $\langle \text{operand} \rangle \rightarrow \mathbf{id}$
6. $\langle \text{operand} \rangle \rightarrow \mathbf{konšt}$

kde

$N = \{ \langle \text{program} \rangle, \langle \text{příkazy} \rangle, \langle \text{příkaz} \rangle, \langle \text{operand} \rangle \}$, $T = \{ \mathbf{id, konšt, \{, \}, ;, +, =} \}$ a počiatočný neterminál je $\langle \text{program} \rangle$. Nájdi deriváčny strom reťazca $\{ \mathbf{id += konšt ; id += id ;} \}$



Derivačný strom:



Redukcia gramatiky

Je daná gramatika $G = \{N, T, P, \acute{S}\}$, kde $N = \{\acute{S}, S, A, B\}$, $T = \{a, b\}$, \acute{S} je počiatočný neterminál a pravidlá:

- $\acute{S} \rightarrow \varepsilon \mid S$
- $S \rightarrow aA \mid aS \mid A \mid abB \mid a$
- $A \rightarrow bA \mid b$
- $B \rightarrow bB$

Vyrobte k nej ekvivalentnú **redukovanú gramatiku**, t.j. odstráňte z nej nadbytočné a nedostupné symboly.



Odstránenie nadbytočných neterminálov

V prvom kroku hľadáme neterminály, ktoré **nie sú nadbytočné**, teda také, ktoré patria do množiny N_T neterminálov, z ktorých je možné teoreticky derivovať nejaký reťazec terminálov:

1. $N_T = \emptyset$ (na začiatku)
2. Keďže $A \rightarrow b$, tak určite $A \in N_T$
3. Keďže $S \rightarrow a$, tak určite $S \in N_T$
4. Keďže $\acute{S} \rightarrow \varepsilon$, tak určite $\acute{S} \in N_T$.
5. Teda $N_T = \{\acute{S}, S, A\}$
6. Ďalej vidíme, že $B \notin N_T$, pretože z B nie je možné odvodiť reťazec terminálov.
7. Preto z gramatiky **odstránime** neterminál B a pravidlá, v ktorých sa vyskytuje.



Dostávame:

- $\acute{S} \rightarrow \varepsilon \mid S$
- $S \rightarrow aA \mid aS \mid A \mid a$
- $A \rightarrow bA \mid b$



Odstránenie nedostupných symbolov

V druhom kroku hľadáme terminály a neterminály, ktoré **sú dostupné**, teda také, ktoré patria do množiny V_D symbolov, ktoré sú dosiahnuteľné z počiatočného neterminálu \hat{S} počas nejakej derivácie:

1. $V_D = \{\hat{S}\}$ (lebo počiatočný neterminál je **vždy** dostupný)
2. Keďže máme pravidlo $\hat{S} \rightarrow S$, aj $S \in V_D$, $V_D = \{\hat{S}, S\}$
3. Keďže máme pravidlo $S \rightarrow aA$, aj $a \in V_D$, $A \in V_D$, $V_D = \{\hat{S}, S, a, A\}$
4. Keďže máme pravidlo $A \rightarrow b$, aj $b \in V_D$, $V_D = \{\hat{S}, S, a, A, b\}$
5. Keďže sme práve zistili, že **každý symbol** gramatiky je dostupný, končíme.
6. Keďže množina V_D obsahuje **všetky symboly** gramatiky, každý symbol je dostupný a teda nič neodstránime.



Gramatika:

- $\acute{S} \rightarrow \varepsilon \mid S$
- $S \rightarrow aA \mid aS \mid A \mid a$
- $A \rightarrow bA \mid b$

je teda **výslednou redukovanou gramatikou**. Každý jej symbol je dostupný a nie nadbytočný. Preto zmazanie **ktoréhokoľvek symbolu** by viedlo k **zмене jazyka**, ktorý gramatika generuje.



Redukcia gramatiky, č. 2

Preved'te uvedenú gramatiku na ekvivalentnú redukovanú gramatiku. Daná gramatika $G = (\{S, A, B, C\}, \{a, b, c\}, P, S)$, pravidlá P :

- $S \rightarrow aAC \mid bB \mid BB$
- $A \rightarrow aAb$
- $B \rightarrow \varepsilon \mid aBb \mid SB$
- $C \rightarrow CC \mid b \mid c$



Množina N_T

1. Na začiatku $N_T = \emptyset$
2. Keďže v pravidlách: $B \rightarrow \varepsilon$, tak $N_T = \{B, \dots$
3. Keďže v pravidlách: $C \rightarrow b$, tak $N_T = \{B, C, \dots$
4. Keď už vieme, že $B \in N_T$ a zároveň v pravidlách $S \rightarrow BB$, tak potom $N_T = \{B, C, S, \dots$
5. A vidíme, že $A \notin N_T$, teda kompletná množina $N_T = \{S, B, C\}$. Neterminál A teda môžeme z gramatiky odstrániť, pretože je nadbytočný.



Dostávame gramatiku:

- $S \rightarrow bB \mid BB$
- $B \rightarrow \varepsilon \mid aBb \mid SB$
- $C \rightarrow CC \mid b \mid c$

Teraz nájdeme množinu V_D .



Množina V_D

1. Na začiatku $V_D = \{S, \dots$
2. Všetky symboly v pravidlách, kde je na ľavej strane S sú taktiež dosiahnuteľné, t.j. $V_D = \{S, b, B, \dots$
3. Keďže už vieme, že aj B je dosiahnuteľný symbol, aj symboly v pravidlách, kde je na ľavej strane B sú dosiahnuteľné, t.j. $V_D = \{S, b, B, a, \dots$
4. Výsledok je teda $V_D = \{S, B, a, b\}$.
5. Symboly C, c teda **nie sú dosiahnuteľné** a môžeme ich zmazať z gramatiky!



Výsledná redukovaná gramatika:

- $S \rightarrow bB \mid BB$
- $B \rightarrow \varepsilon \mid aBb \mid SB$



POZOR!!!

Ak by sme v danej gramatike:

- $S \rightarrow aAC \mid bB \mid BB$
- $A \rightarrow aAb$
- $B \rightarrow \varepsilon \mid aBb \mid SB$
- $C \rightarrow CC \mid b \mid c$

najprv odstránili nedostupné symboly (množina V_D) a potom nadbytočné neterminály (množina N_T), t.j. odstránenie symbolov vykonáme v **opačnom poradí**, **NEDOSTANEME** redukovanú gramatiku!!!

Pre túto gramatiku totiž $V_D = \{S, a, A, C, b, B\}$, t.j. všetky symboly sú dosiahnuteľné, čiže by sme neodstránili nič a následne $N_T = \{S, B, C\}$, čiže by sme dostali gramatiku

- $S \rightarrow bB \mid BB$
- $B \rightarrow \varepsilon \mid aBb \mid SB$
- $C \rightarrow CC \mid b \mid c$

ktorá **nie je redukovaná**. Preto na poradí odstraňovania symbolov **záleží!**



Gramatika na úpravu č. 3

Upravte uvedenú gramatiku na redukovanú gramatiku:

- $S \rightarrow Aa \mid Bb$
- $A \rightarrow C \mid DaD$
- $B \rightarrow D \mid C$
- $C \rightarrow S \mid a \mid \varepsilon$
- $D \rightarrow S \mid b$
- $E \rightarrow S \mid c \mid \varepsilon$



Množina N_T

- $N_T = \emptyset$
- Keďže v pravidlách $C \rightarrow a$, tak $C \in N_T$
- Keďže v pravidlách $D \rightarrow b$, tak $D \in N_T$
- Keďže v pravidlách $E \rightarrow c$, tak $E \in N_T$
- Keďže v pravidlách $A \rightarrow c$ a $C \in N_T$, tak aj $A \in N_T$
- Keďže v pravidlách $B \rightarrow D$ a $D \in N_T$, tak aj $B \in N_T$
- Keďže v pravidlách $S \rightarrow Aa$ a $A \in N_T$, tak aj $S \in N_T$
- Teda $N_T = \{S, A, B, C, D, E\}$ a žiaden neterminál v tomto kroku neodstránime.



Množina V_D

- $V_D = \{S\}$
- Do množiny V_D pribudnú všetky symboly dosiahnuteľné z neterminálu S , t.j. z pravidiel, kde je S na ľavej strane, $V_D = \{S, A, a, B, b\}$.
- Do množiny V_D pribudnú všetky symboly dosiahnuteľné z neterminálu A , t.j. z pravidiel, kde je A na ľavej strane, $V_D = \{S, A, a, B, b, C, D\}$.
- Do množiny V_D pribudnú všetky symboly dosiahnuteľné z neterminálu B , t.j. z pravidiel, kde je A na ľavej strane, $V_D = \{S, A, a, B, b, C, D\}$.
- Do množiny V_D pribudnú všetky symboly dosiahnuteľné z neterminálu C , t.j. z pravidiel, kde je A na ľavej strane, $V_D = \{S, A, a, B, b, C, D\}$.
- Do množiny V_D pribudnú všetky symboly dosiahnuteľné z neterminálu D , t.j. z pravidiel, kde je A na ľavej strane, $V_D = \{S, A, a, B, b, C, D\}$.
- Vidíme, že $V_D = \{S, A, B, C, D, a, b\}$. Teda symboly E, c sú nedosiahnuteľné a môžeme ich odstrániť.



Redukovaná gramatika

Redukovaná gramatika:

- $S \rightarrow Aa \mid Bb$
- $A \rightarrow C \mid DaD$
- $B \rightarrow D \mid C$
- $C \rightarrow S \mid a \mid \varepsilon$
- $D \rightarrow S \mid b$



Gramatika na úpravu č. 4

Upravte uvedenú gramatiku na redukovanú gramatiku:

- $S \rightarrow ABC$
- $A \rightarrow abA \mid BbC$
- $B \rightarrow Aa \mid BB$
- $C \rightarrow S \mid AB \mid \varepsilon$



Množina N_T

- $N_T = \emptyset$
- Keďže v pravidlách $C \rightarrow \varepsilon$, tak $C \in N_T$
- Keďže v pravidlách nie sú žiadne ďalšie pravidlá, ktoré by mali napravo alebo reťazec nad terminálmi, reťazec zložený z terminálov a neterminálu C , algoritmus končí a výsledná množina $N_T = \{C\}$
- Odstránili by sme teda neterminály S, A, B , teda **aj počiatočný neterminál!!!**
- Ak nastáva situácia, že **počiatočný neterminál nepatrí do množiny N_T** , tak potom **gramatika generuje len prázdny jazyk** a nemá zmysel pokračovať v jej redukcii!

